Volume 5 Number 12 December 1978 Nucleic Acids Research

Nucleotide sequence of bacteriophage fd DNA

E.Beck*, R.Sommer, E.A.Auerswald, Ch.Kurz, B.Zink, G.Osterburg†, and H.Schaller
Microbiology, Univ. Heidelberg, Im Neuenheimer Feld 230, †Inst. Documentation, Information and Statistic, German Cancer Research Cent.,Im Neuenheimer Feld 280, 6900 Heidelberg, GFR and

K.Sugimoto, H.Sugisaki, T.Okamoto and M.Takanami.
Institute for Chemical Research, Kyoto University, Uji, Kyoto-Fu 611, Japan

Received 6 November 1978

ABSTRACT

The sequence of the 6408 nucleotides of bacteriophage fd DNA has been determined. This allows to deduce the exact organisation of the filamentous phage genome and provides easy access to DNA segments of known structure and function.

INTRODUCTION

Small DNA viruses depend during their life cycle largely on host functions and are therefore preferred model systems for the analysis of the organisation, expression and replication of the more complex host genomes. To analyse viral genomes at the nucleotide level has become technically possible with the development of new rapid DNA sequencing techniques[1,2,3]. Complete nucleotide sequences have been reported so far for coli phage $\phi X174$[4,5] and Simian Virus SV40[6,7]. Here we report the sequence of bacteriophage fd DNA, strain 478 (Heidelberg).

Phage fd[8] along with f1 and M13 belongs to a group of closely related filamentous, male-specific coli phages (for reviews see ref. 9, 10). Its genome is a single-stranded circular DNA of about 6000 nucleotides which is converted to a double-stranded form in the infected cell. Eight genes have been ordered by combined genetic and biochemical analysis within the phage genome. Its detailed organisation remained, however, relatively uncertain due to the lack of protein data for most gene products. Furthermore, analysis on the nucleotide level had concentrated mainly on DNA segments with regulatory functions[10,11].

© Information Retrieval Limited 1 Falconberg Court London W1V 5FG England

We have previously reported a preliminary nucleotide sequence of fd DNA ([11], and personal communications). The aim of this publication is the rapid communication of the final sequence. A more detailed account containing the experimental evidence will be published elsewhere.

RESULTS AND DISCUSSION

Restriction nucleases and cleavage maps. The enzymes used, their recognition sequences and the position of cleavage sites confirmed or newly established during this work are presented in Fig. 1. All cleavage sites shown have also been identified by DNA sequencing the ends of the respective restriction fragments. With one exception, all parts of double-stranded fd can be fragmented by digestion with several of these enzymes into pieces of less than 200 base-pairs.

DNA sequencing. The chemical method of Maxam and Gilbert[2] was used which allowed us to read sequences up to 150 (occasionally up to 220) nucleotides. Sequences obtained were stored and processed in a computer (G. Osterburg and R. Sommer, to be published) to yield the composite sequence of 6408 nucleotides presented in Fig. 2. About 75 % of this sequence was determined from both DNA strands in fd 478. Almost all of the missing 25 % have also been sequenced in the second strand, but in the closely related phage f1. Further information was obtained for about 1000 nucleotides by RNA sequencing[12] and for about 600 nucleotides by the plus/minus method of Sanger and Coulson[1]. About 10 % of the fd sequence were also established as recognition sequences for restriction nucleases at known cleavage sites (Fig. 1 and unpublished results).

Nucleotide sequence. According to Fig. 2 fd DNA is composed of 6408 nucleotides (1578A, 2210T, 1325G, 1295C) corresponding to a molecular weight of 2.12×10^6 daltons) (sodium salt). The sequence differs from that reported earlier[11] mainly by an insert of 18 nucleotides in the

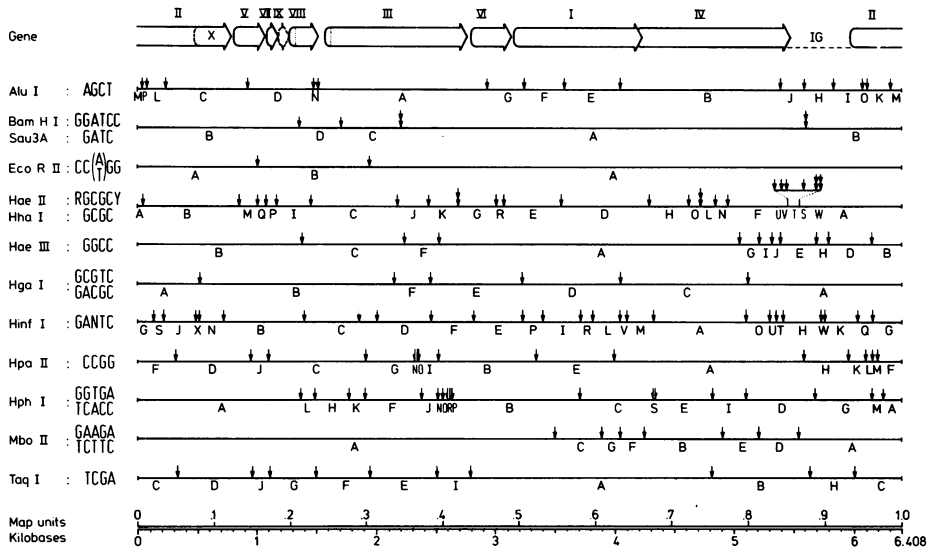


Fig. 1: Fragment maps of restriction nucleases used in the sequence analysis of fd DNA, strain 478. The known maps for HpaII, HgaI, HaeII (HinHI), HaeIII, Alu[10,11] were confirmed and refined. Maps for HhaI, HinfI, TaqI, BamHI, Sau3A (DpnI, MboI), EcoRII, MboII, and HphI were newly established (E.A. Auerswald et al., M. Takanami et al., both unpublished). The first nucleotide of the recognition sites for the various restriction nucleases are listed below. An additional Hinf site has been detected in fragment HinfC (position 1858) in the DNA from fd ATCC (M. Takanami, unpublished). The circular phage DNA is opened at the unique HindII (HpaI) cleavage site. The map includes the positions and the orientation of the phage genes. IG is the intergenic space.

AluI	AGCT	39	63	229	934	1488	1517	2963	3277	3613	4097	5427	5631
		5888	6108	6135	6336								
BamHI	GGATCC	2220	5645										
Sau3A	GATC	1382	1714	2221	5646								
EcoRII	CCTGG	1014	1966										
HaeII	RGCGCY	2710	4743	5560	5568								
HhaI	GCGC	44	873	1011	1085	1177	1470	2195	2467	2711	3040	3096	3599
		4313	4642	4744	4886	4996	5491	5504	5513	5535	5561	5569	
HaeIII	GGCC	1396	2245	2554	5082	5240	5346	5415	5726	5829	6181		
HgaI	GACGC	526	2164	2479	3238								
	GCGTC	4084	5159										
HinfI	GANTC	136	216	490	511	723	1403	2011	2497	2845	3259	3419	3743
		3839	4073	4118	4350	5121	5330	5376	5439	5767	5789	6043	6199
HpaII	CCGG	314	966	1095	1924	2378	2390	2396	2552	3371	4019	5615	5996
		6119	6179	6221									
HphI	GGTGA	1376	1774	1909	2398	2542	2581	2620	2626	3740	4347	4848	5118
		5707	6163										
	TCACC	1503	2635	4365	6189	6286							
MboII	GAAGA	3913											
	TCTTC	3529	4076	4272	4938	5256	5588						
TaqI	TCGA	336	988	1127	1508	1949	2528	2815	4834	5684	6041		

repetitive sequence around position 2380. Except for a G → A transition in position 1859 the identical sequence was obtained in 2000 nucleotides from another fd strain (ATCC).

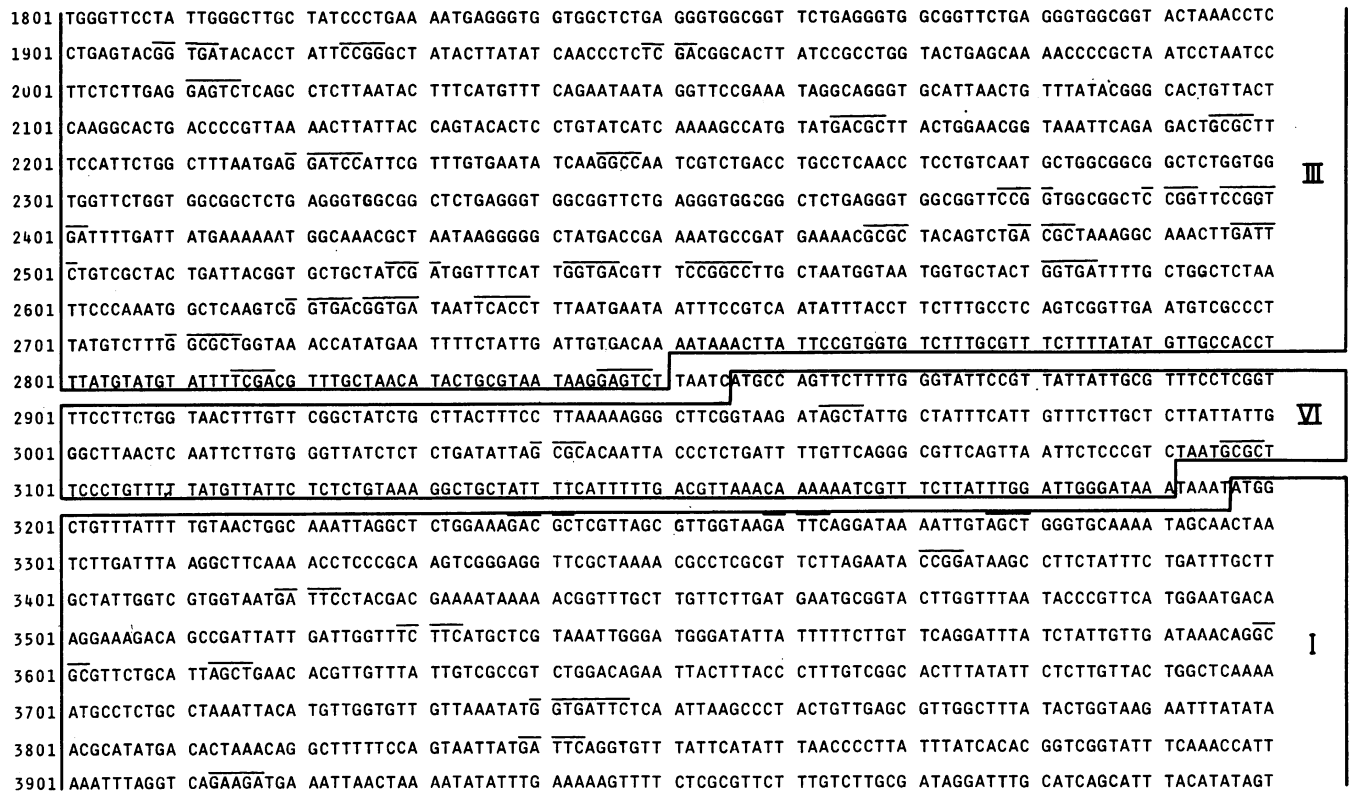
The nucleotide sequence of the related phage f1 has been determined to about 90 % (E. Beck, unpublished). It differs from the fd sequence by deletion of a single nucleotide (position 3195) and by about 160 base changes. Except for seven, these are all silent mutations which do not alter the amino acid sequence of the fd gene products.

Genome organisation. By analysing the fd DNA sequence for continuous translational reading frames - combined with the information obtained from the sequence of amber mutations in f1 and M13 (E. Beck, unpublished; J. Schoenmakers, personal communication) and from the silent base changes in f1 - allows to deduce the exact sizes and positions of the eight known gene products and of known regulatory signals. The DNA sequence predicts the amino acid sequences of known and unknown gene products, and the existence of a new gene (gene IX) in the intergenic space between genes VII and VIII[11].

According to our analysis (Fig. 2) the overall organisation of the filamentous phage genome differs markedly from that of icosahedral single-stranded DNA phages, like $\phi X174$[4]: Although genes are generally closely spaced there is only one single short overlap of genes in different reading frames (at the junction of genes I and IV). In addition there is an intergenic region (IG) of 508 nucleotides which harbours the origins of DNA replication[13,14]. Recent experiments show that this space can be further expanded by insertion of foreign DNA[16].

Applications. fd DNA is accessible in high yields in both its single-stranded and double-stranded form[10]. The knowledge of its nucleotide sequence and of the map positions of a great number of restriction sites provides therefore easy access to well defined DNA molecules which can be used in different investigations on DNA structure


5782 AATAGTG GACTCTTGTT CCAAACTGGA
5809 ACAACACTCA CAACTAACTC GGCCTATTCT TTTGATTTAT AAGGATTTTT GTCATTTTCT GCTTACTGGT TAAAAAATAA GCTGATTTAA CAAATATTTA
5909 ACGCGAAATT TAACAAAACA TTAACGTTTA CAATTTAAAT ATTTGCTTAT ACAATCATCC TGTTTTTGGG GCTTTTCTGA TTATCAACCG GGGTACATAT
6009 GATTGACATG CTAGTTTTAC GATTACCGTT CATCGATTCT CTTGTTTGCT CCAGACTTTC AGGTAATGAC CTGATAGCCT TTGTAGACCT CTCAAAAATA
6109 GCTACCCTCT CCGGCATGAA TTTATCAGCT AGAACGGTTG AATATCATAT TGACGGTGAT TTGACTGTCT CCGGCCTTTC TCACCCGTTT GAATCTTTGC
6209 CTACTCATTA CTCCGGCATT GCATTTAAAA TATATGAGGG TTCTAAAAAT TTTTATCCCT GCGTTGAAAT TAAGGCTTCA CCAGCAAAAG TATTACAGGG
6309 TCATAATGTT TTTGGTACAA CCGATTTAGC TTTATGCTCT GAGGCTTTAT TGCTTAATTT TGCTAACTCT CTGCCTTGCT TGTACGATTT ATTGGATGTT
1 AACGCTACTA CCATTAGTAG AATTGATGCC ACCTTTTCAG CTCGCGCCCC AAATGAAAAT ATAGCTAAAC AGGTTATTGA CCATTTGCGA AATGTATCTA
101 ATGGTCAAAC TAAATCTACT CGTTCGCAGA ATTGGGAATC AACTGTTACA TGGAATGAAA CTTCCAGACA CCGTACTTTA GTTGCATATT TAAAACATGT
201 TGAACTACAG CACCAGATTC AGCAATTAAG CTCTAAGCCA TCCGCAAAAA TGACCTCTTA TCAAAAGGAG CAATTAAAGG TACTGTCTAA TCCTGACCTG
301 TTGGAATTTG CTTCCGGTCT GGTTCGCTTT GAGGCTCGAA TTGAAACGCG ATATTTGAAG TCTTTCGGGC TTCCTCTTAA TCTTTTTGAT GCAATTCGCT
401 TTGCTTCTGA CTATAATAGA CAGGGTAAAG ACCTGATTTT TGATTTATGG TCATTCTCGT TTTCTGAACT GTTTAAAGCA TTTGAGGGGG ATTCAATGAA
501 TATTTATGAC GATTCCGCAG TATTGGACGC TATCCAGTCT AAACATTTTA CAATTACCCC CTCTGGCAAA ACTTCCTTTG CAAAAGCCTC TCGCTATTTT
601 GGTTTCTATC GTCGTCTGGT TAATGAGGGT TATGATAGTG TTGCTCTTAC CATGCCTCGT AATTCCTTTT GGCGTTATGT ATCTGCATTA GTTGAGTGTG
701 GTATTCCTAA ATCTCAATTG ATGAATCTTT CCACCTGTAA TAATGTTGTT CCGTTAGTTC GTTTTATTAA CGTAGATTTT TCCTCCCAAC GTCCTGACTG
801 GTATAATGAG CCAGTTCTTA AAATCGCATA AGGTAATTCA AAATGATTAA AGTTGAAATT AAACCGTCTC AAGCGCAATT TACTACCCGT TCTGGTGTTT
901 CTCGTCAGGG CAAGCCTTAT TCACTGAATG AGCAGCTTTG TTACGTTGAT TTGGGTAATG AATATCCGGT GCTTGTCAAG ATTACTCTCG ACGAAGGTCA
1001 GCCAGCGTAT GCGCCTGGTC TGTACACCGT GCATCTGTCC TCGTTCAAAG TTGGTCAGTT CGGTTCTCTT ATGATTGACC GTCTGCGCCT CGTTCCGGCT
1101 AAGTAACATG GAGCAGGTCG CGGATTTCGA CACAATTTAT CAGGCGATGA TACAAATCTC CGTTGTACTT TGTTTCGCGC TTGGTATAAT CGCTGGGGGT
1201 CAAAGATGAG TGTTTTAGTG TATTCTTTCG CCTCTTTCGT TTTAGGTTGG TGCCTTCGTA GTGGCATTAC GTATTTTACC CGTTTAATGG AAACTTCCTC
1301 ATGAAAAAGT CTTTAGTCCT CAAAGCCTCC GTAGCCGTTG CTACCCTCGT TCCGATGCTG TCTTTCGCTG CTGAGGGTGA CGATCCCGCA AAAGCGGCCT
1401 TTGACTCCCT GCAAGCCTCA GCGACCGAAT ATATCGGTTA TGCGTGGGCG ATGGTTGTTG TCATTGTCGG CGCAACTATC GGTATCAAGC TGTTTAAGAA
1501 ATTCACCTCG AAAGCAAGGT GATAAACCGA TACAATTAAA GGCTCCTTTT GGAGCCTTTT TTTTTGGAGA TTTTCAACGT GAAAAAATTA TTATTCGCAA
1601 TTCCTTTAGT TGTTCCTTTC TATTCTCACT CCGCTGAAAC TGTTGAAAGT TGTTTAGCAA AACCTCATAC AGAAAATTCA TTTACTAACG TCTGGAAAGA
1701 CGACAAAACT TTAGATCGTT ACGCTAACTA TGAGGGCTGT CTGTGGAATG CTACAGGCGT TGTGGTTTGT ACTGGTGACG AAACTCAGTG TTACGGTACA
IG
II
X
V
VII
IX
VIII
III


1801 TGGGTTCCTA TTGGGCTTGC TATCCCTGAA AATGAGGGTG GTGGCTCTGA GGGTGGCGGT TCTGAGGGTG GCGGTTCTGA GGGTGGCGGT ACTAAACCTC
1901 CTGAGTACGG TGATACACCT ATTCCGGGCT ATACTTATAT CAACCCTCTC GACGGCACTT ATCCGCCTGG TACTGAGCAA AACCCCGCTA ATCCTAATCC
2001 TTCTCTTGAG GAGTCTCAGC CTCTTAATAC TTTCATGTTT CAGAATAATA GGTTCCGAAA TAGGCAGGGT GCATTAACTG TTTATACGGG CACTGTTACT
2101 CAAGGCACTG ACCCCGTTAA AACTTATTAC CAGTACACTC CTGTATCATC AAAAGCCATG TATGACGCTT ACTGGAACGG TAAATTCAGA GACTGCGCTT
2201 TCCATTCTGG CTTTAATGAG GATCCATTCG TTTGTGAATA TCAAGGCCAA TCGTCTGACC TGCCTCAACC TCCTGTCAAT GCTGGCGGCG GCTCTGGTGG
2301 TGGTTCTGGT GGCGGCTCTG AGGGTGGCGG CTCTGAGGGT GGCGGTTCTG AGGGTGGCGG CTCTGAGGGT GGCGGTTCCG GTGGCGGCTC CGGTTCCGGT
2401 GATTTTGATT ATGAAAAAAT GGCAAACGCT AATAAGGGGG CTATGACCGA AAATGCCGAT GAAAACGCGC TACAGTCTGA CGCTAAAGGC AAACTTGATT
2501 CTGTCGCTAC TGATTACGGT GCTGCTATCG ATGGTTTCAT TGGTGACGTT TCCGGCCTTG CTAATGGTAA TGGTGCTACT GGTGATTTTG CTGGCTCTAA
2601 TTCCCAAATG GCTCAAGTCG GTGACGGTGA TAATTCACCT TTAATGAATA ATTTCCGTCA ATATTTACCT TCTTTGCCTC AGTCGGTTGA ATGTCGCCCT
2701 TATGTCTTTG GCGCTGGTAA ACCATATGAA TTTTCTATTG ATTGTGACAA AATAAACTTA TTCCGTGGTG TCTTTGCGTT TCTTTTATAT GTTGCCACCT
2801 TTATGTATGT ATTTTCGACG TTTGCTAACA TACTGCGTAA TAAGGAGTCT TAATCATGCC AGTTCTTTTG GGTATTCCGT TATTATTGCG TTTCCTCGGT
2901 TTCCTTCTGG TAACTTTGTT CGGCTATCTG CTTACTTTCC TTAAAAAGGG CTTCGGTAAG ATAGCTATTG CTATTTCATT GTTTCTTGCT CTTATTATTG
3001 GGCTTAACTC AATTCTTGTG GGTTATCTCT CTGATATTAG CGCACAATTA CCCTCTGATT TTGTTCAGGG CGTTCAGTTA ATTCTCCCGT CTAATGCGCT
3101 TCCCTGTTTJ TATGTTATTC TCTCTGTAAA GGCTGCTATT TTCATTTTTG ACGTTAAACA AAAAATCGTT TCTTATTTGG ATTGGGATAA ATAAATATGG
3201 CTGTTTATTT TGTAACTGGC AAATTAGGCT CTGGAAAGAC GCTCGTTAGC GTTGGTAAGA TTCAGGATAA AATTGTAGCT GGGTGCAAAA TAGCAACTAA
3301 TCTTGATTTA AGGCTTCAAA ACCTCCCGCA AGTCGGGAGG TTCGCTAAAA CGCCTCGCGT TCTTAGAATA CCGGATAAGC CTTCTATTTC TGATTTGCTT
3401 GCTATTGGTC GTGGTAATGA TTCCTACGAC GAAAATAAAA ACGGTTTGCT TGTTCTTGAT GAATGCGGTA CTTGGTTTAA TACCCGTTCA TGGAATGACA
3501 AGGAAAGACA GCCGATTATT GATTGGTTTC TTCATGCTCG TAAATTGGGA TGGGATATTA TTTTTCTTGT TCAGGATTTA TCTATTGTTG ATAAACAGGC
3601 GCGTTCTGCA TTAGCTGAAC ACGTTGTTTA TTGTCGCCGT CTGGACAGAA TTACTTTACC CTTTGTCGGC ACTTTATATT CTCTTGTTAC TGGCTCAAAA
3701 ATGCCTCTGC CTAAATTACA TGTTGGTGTT GTTAAATATG GTGATTCTCA ATTAAGCCCT ACTGTTGAGC GTTGGCTTTA TACTGGTAAG AATTTATATA
3801 ACGCATATGA CACTAAACAG GCTTTTTCCA GTAATTATGA TTCAGGTGTT TATTCATATT TAACCCCTTA TTTATCACAC GGTCGGTATT TCAAACCATT
3901 AAATTTAGGT CAGAAGATGA AATTAACTAA AATATATTTG AAAAAGTTTT CTCGCGTTCT TTGTCTTGCG ATAGGATTTG CATCAGCATT TACATATAGT
III
VI
I

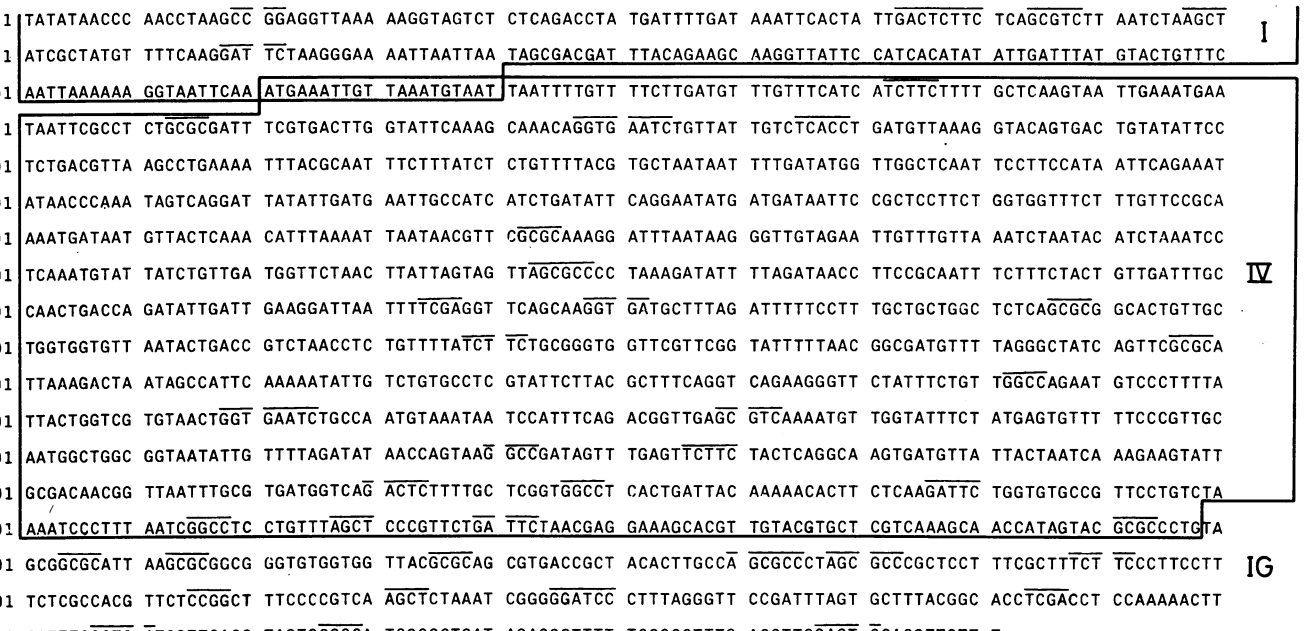


Fig. 2: Nucleotide sequence of bacteriophage fd. The viral DNA single-strand is shown in 5' → 3' polarity. The circular DNA has been opened at the position of the origin of viral replication[13 14]. Numbering of nucleotides starts at the unique HindII (HpaI) cleavage site. Genes are boxed, recognition sites for the restriction nucleases shown in Fig. 1 are overlined. The sequence is available on request on magnetic tape.

and function. For example they have been used as size markers in their intact or restricted form, for the search for recognition sequences of restriction nucleases[17], in the site-specific modification of the fd genome for use as a cloning vehicle[16], for the isolation and the cloning of regulatory signals from fd DNA, for the analysis of integration and loss of transposon Tn-5 ([16,14], E.A. Auerswald, to be published), and for the correlation of thermal denaturation profiles of DNA molecules with their nucleotide sequence[18].

ACKNOWLEDGEMENT

This work was supported by a grant from the Deutsche Forschungsgemeinschaft.

*Present address: The University of British Columbia, Faculty of Medicine, Department of Biochemistry, 2075 Westbrook Place, Vancouver, B.C. Canada V6T 1W5.

REFERENCES

1 Sanger, F. and Coulson, A.R. (1975) J. Mol. Biol. 94, 441 - 448

2 Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560 - 564

3 Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463 - 5467

4 Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchison III, C.A., Slocombe, P.M., and Smith, M. (1977) Nature 265, 687 - 695

5 Sanger, F. and Coulson, A.R. (1978) FEBS Letters 87, 107 - 110

6 Fiers, W., Contreras, W.R., Haegeman, G., Rogiers, R., van de Voorde, A., van Heuverswyn, H., van Herreweghe, J., Volckaert, G., and Ysebaert, M. (1978) Nature 273 113 - 120

7 Reddy, V.B., Thimmappaya, B., Dhar, R., Subramanian, K.N., Zain, B.S., Pan, J., Ghosh, P.K., Celma, M.L., Weissman, S.M. (1978) Science 200, 494 - 502

8 Hoffmann-Berling, H., Marvin, D.A., and Dürwald, H. (1963) in Zeitschrift für Naturforschung Bd. 18b, Heft 11, 876 - 898

9 Marvin, D.A. and Hohn, B. (1969) Bacteriol. Rev. 33, 172 - 209

10 Ray, D.S. (1977) in Comprehensive Virol. 7, 105 - 178 eds. Fraenkel-Conrat, H. and Wagner, R.R., Plenum Publ. Corp. N.Y. 10011 USA

11 Schaller, H., Beck, E., and Takanami, M. (1978) in Single Stranded DNA Phages, eds. Denhardt, D.T., Dressler, D., and Ray, D.S., Cold Spring Harbor Lab., Cold Spring Harbor, N.Y. 11724 USA

12 Sugimoto, K., Sugisaki, H., Okamoto, T., and Takanami, M. (1977) J. Mol. Biol. 111, 487 - 507, and unpublished results

13 Horiuchi, K. and Zinder, N.D. (1976) Proc. Natl. Acad. Sci. USA 73, 2341 - 2345

14 Schaller, H. (1978) in Symp. on Quantitative Biol. Vol. XLIII, Cold Spring Harbor Lab., Cold Spring Harbor, N.Y. 11724, USA

15 Meyer, T., Geider, K., Kurz, Ch., and Schaller, H. (submitted for publication)

16 Herrmann, R., Neugebauer, K., Zentgraf, H., and Schaller, H. (1978) in Single Stranded DNA Phages, eds. Denhardt, D.T., Dressler, D., and Ray, D.S., Cold Spring Harbor Lab., Cold Spring Harbor, N.Y. 11724 USA

17 Sommer, R. and Schaller, H. (1979) Molec. gen. Genet. (in press)

18 Vologodskii, A.V. and Frank-Kamenetskii, M.D. (1978) Nucleic Acids Res. 5, 2547 - 2556